Remote Sensing of Vegetation Characteristics in Support of Ecological Modelling

Roelofsen, H.D.

2014

document version
Publisher's PDF, also known as Version of record

Chapter 5. Mapping *a priori* defined plant associations using remotely sensed vegetation characteristics

Hans D. Roelofsen[1]

Lammert Kooistra[2]

Peter M. van Bodegom[3]

Jochem Verrelst[4]

Johan Krol[5]

Jan-Philip M. Witte[1, 3]

1. KWR Watercycle Research Institute, Nieuwegein, the Netherlands
2. Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Wageningen, the Netherlands
3. Department of Ecological Sciences, Sub Department Systems Ecology, VU University, Amsterdam, the Netherlands
4. Image Processing Laboratory, University of Valencia, Valencia, Spain
5. Nature Centre Ameland, Nes, the Netherlands

Remote Sensing of Environment, (2014), 140, 639-651

5.1. Abstract

Incorporation of a priori defined plant associations into remote sensing products is a major challenge that has only recently been confronted by the remote sensing community. We present an approach to map the spatial distribution of such associations by using plant indicator values (IVs) for salinity, moisture and nutrients as an inter-mediate between spectral reflectance and association occurrences. For a 12 km² study site in the Netherlands, the relations between observed IVs at local vegetation plots and visible and near-infrared (VNIR) and short-wave infrared (SWIR) airborne reflectance data were modelled using Gaussian Process Regression (GPR) (R^2 0.73, 0.64 and 0.76 for salinity, moisture and nutrients, respectively). These relations were applied to map IVs for the complete study site. Association occurrence probabilities were modelled as function of IVs using a large database of vegetation plots with known association and IVs. Using the mapped IVs, we calculated occurrence probabilities of 19 associations for each pixel, resulting in both a crisp association map with the most likely occurring association per pixel, as well as occurrence probability maps per association. Association occurrence predictions were assessed by a local vegetation expert, which revealed that the occurrences of associations situated at frequently predicted indicator value combinations were over predicted. This seems primarily due to biases in the GPR predicted IVs, resulting in associations with envelopes located in extreme ends of IVs being scarcely predicted. Although the results of this particular study were not fully satisfactory, the method potentially offers several ad-vantages compared to current vegetation classification techniques, like site-independent calibration of association probabilities, site-independent selection of associations and the provision of IV maps and occurrence probabilities per association. If the prediction of IVs can be improved, this method may thus provide a viable roadmap to bring a priori defined plant associations into the domain of remote sensing.

5.2. Introduction

Mapping the characteristics and extent of natural vegetation using remote sensing has been the subject of many studies in recent years (see for a review: (Ustin & Gamon, 2010)). A common challenge is to simplify and generalise the complex interactions in natural vegetation so that meaningful information on abiotic and biotic conditions can be derived (Janssen, 2004; Küchler & Zonneveld, 1988; Küchler, 1984; Sanders et al., 2004). A typical method for this is to define vegetation units for the specific site and subsequently delineate those using the remote sensing data and a classification technique (Belluco et al., 2006; Kokaly et al., 2003; Oldeland et al., 2010; Thomas et al., 2003). The vegetation units are often based on vegetation properties that are typically well discernible by remote sensing techniques, such as vegetation structure, bio-chemistry and phenology (Ustin & Gamon, 2010). Other approaches ignore the concept of crisp vegetation units and visualise the continuous properties of vegetation, such as position on floristic gradients (Feilhauer et al., 2011), forest diversity (Feilhauer & Schmidtlein, 2009), plant strategies (Schmidtlein et al., 2011) or fractional cover of vegetation and soil (Asner & Heidebrecht, 2002). Other studies

aim to isolate a particular feature of interest, such as nonnative species (Underwood et al., 2003) or invasive woody species (Hantson et al., 2012). Discrimination and identification of individual tree or shrub species is currently feasible in some ecosystems (Cho et al., 2010; Clark et al., 2005; Dennison & Roberts, 2003; Roth et al., 2012; Xiao et al., 2004) and may benefit from combining LiDAR and hyperspectral data (Asner & Martin, 2009).

The capacities of remote sensing might be more comprehensively exploited when remote sensing products reach potential end users such as nature managers and policy makers. However, such profession-al nature management and conservation agencies often employ a specific–and different from above–definition of vegetation units to describe their assets. Instead, these vegetation units are commonly created using a phytosociological approach, where species abundance data are used in cluster analysis and ordination to create discrete vegetation units (Dengler et al., 2008) that are differentiated by the presence or absence of diagnostic species (Verrelst et al., 2009). Current sensors are not suited to detect diagnostic species, all the more since they usually are sparse and have a low abundance (Verrelst et al., 2009). Therefore, incorporation of phytosociological vegetation units has only recently been confronted by the remote sensing community, but not without complications. For example, Schmidt et al. (2004) found that phytosociological units could only be successfully mapped after consulting expert knowledge in addition to hyperspectral and LiDAR data. Verrelst et al. (2009) chose to map clusters of phytosociological vegetation units because the individual vegetation units could not be differentiated directly. Given such difficulties, a comprehensive method that translates remote sensing data to vegetation units that are defined a priori and based on species composition, like phytosociological units, is desired.

We explore a method that uses vegetation characteristics such as plant traits (i.e., morphological, physiological or phenological features measurable at the individual plant level, (Violle et al., 2007)) or indicator values (IVs, (Diekmann, 2002), see Section 2.1 for definition), as an intermediate between spectral reflectance and vegetation units. Such vegetation characteristics relate directly to plant biochemical, biophysical and phenological properties; properties that are also observable with remote sensing techniques (Asner et al., 2011; Asner & Martin, 2008; Schmidtlein, 2005; Ustin & Gamon, 2010; Verrelst, Muñoz, et al., 2012). In addition, it has been demonstrated that variation in vegetation characteristics is constrained by local environmental conditions and that plants themselves have limited ability in adapting their characteristic values upon a change in environmental conditions (Ackerly & Cornwell, 2007; de Bello et al., 2009). Consequently, different vegetation units may be distributed along an n-dimensional vegetation characteristic space. This holds true for a priori defined vegetation units; their occurrence probability can be calculated for a given set of plant traits (Douma, Aerts, et al., 2012) or IVs (Witte et al., 2007). This means that vegetation characteristics as predicted by remote sensing may be used to calculate vegetation unit occurrence probability.

This paper evaluates this hypothesis by using plant IVs to differentiate between a priori defined vegetation units that are often used among ecologists in the Netherlands. Two key concepts, IVs and the vegetation units, are described in detail first. Following this, a case study is presented where 19 representative vegetation units have been mapped. The

resulting maps were validated by expert judgement and by comparison with a pre-existing vegetation map.

5.3. Method & materials

5.3.1. General methodology

A well-known method for determining a priori defined vegetation units is the school of Braun-Blanquet, also called 'phytosociology'. Here, discrete vegetation units, collectively called syntaxa, are clustered hierarchically into (from low to high ranks) associations, alliances, orders and classes. The integral hierarchical tree of these syntaxa is referred to as a syntaxonomical system (Weber et al., 2000). The association is the basic unit and defined as 'a plant community of definite floristic composition which presents a uniform physiognomy and which grows in uniform habitat conditions' (Weber et al., 2000). Plant associations are defined by the presence of diagnostic species (in particular: 'character' species and 'differential' species). As a showcase for the proposed method, we mapped plant associations according to the Dutch standard work of phytosociology 'De Vegetatie van Nederland' (DVN, the vegetation of the Netherlands, (Schaminée, Stortelder, et al., 1995)). This system defines 228 associations which are grouped into 89 alliances, 58 orders and 43 classes. In addition to the syntaxon names, the syntaxa in DVN are labelled with a 6 digit code from which their position in the hierarchical structure can be derived; the first two numbers indicate the class, the third and fourth characters indicate the order and alliance respectively, while the final two numbers determine the association. In this paper, the relevant associations are introduced once with their name and code and subsequently identified with their code only.

The method investigated in the paper pivots on the concept of using vegetation characteristics as an intermediate between spectral reflectance and occurrence of plant associations. In this case study, we use IVs as vegetation characteristics that couple the remote sensing data to associations. Originally introduced by Ellenberg et al. (1991), IVs use rankings of plant species occurrences to identify the most common occurrence of a species along a normalised environmental gradient (Schmidtlein, 2005). IVs are widely used in plant and systems ecology, but their use is also criticised due to supposed subjectivity and potential circular reasoning (Diekmann, 2002; Klaus et al., 2012; Schaffers & Sykora, 2000; Zelený & Schaffers, 2012). It is important to note that an IV is not a physical quantity that can be measured on a plant, but rather an artificially constructed property of a plant species. Another important aspect of our approach is that we did not use Ellenberg IVs. Instead, we employed IVs that are designed specifically for the Netherlands and that have non integer values for individual species. This continuous representation is based on direct field observations (instead of expert knowledge) and allows for more reliable calculation of vegetation plot mean IVs. We used an already existing list of IVs per plant species that was compiled earlier; refer to Witte et al. (2007) for a detailed explanation on how this list has been compiled.

The general methodology is represented in Fig. 5.1. Firstly, in order to map IVs for the whole site, observed IVs of local vegetation plots were related to top of canopy reflectance values derived from airborne optical remote sensing imagery using Gaussian Process Regression (GPR) (step 1, Fig. 5.1). GPR model accuracy was assessed using an independent validation set of local vegetation plots. Secondly, Gaussian Mixture Density Modelling (GMDM) was used to define Probability Density Functions (PDFs) from a national database of vegetation plots (step 2, Fig. 5.1). Each PDF describes the Bayesian probability of association occurrence at a given combination of IVs. Finally, based on predicted IVs (step 1, Fig. 5.1) and the PDFs of the associations (step 2, Fig. 5.1), Bayesian occurrence probabilities were calculated for each association at each pixel (step 3, Fig. 5.1), visualising the occurrence probability per association. In addition, the association corresponding to the highest occurrence probability was assigned to each pixel, creating a crisp classification of associations (step 3, Fig. 5.1). The certainty of the predicted association was represented as the maximum occurrence probability per pixel. A local vegetation expert assessed the accuracy of the predicted association occurrence. In addition, we compared the predicted distribution patterns per association with a preexisting vegetation map of the study site.

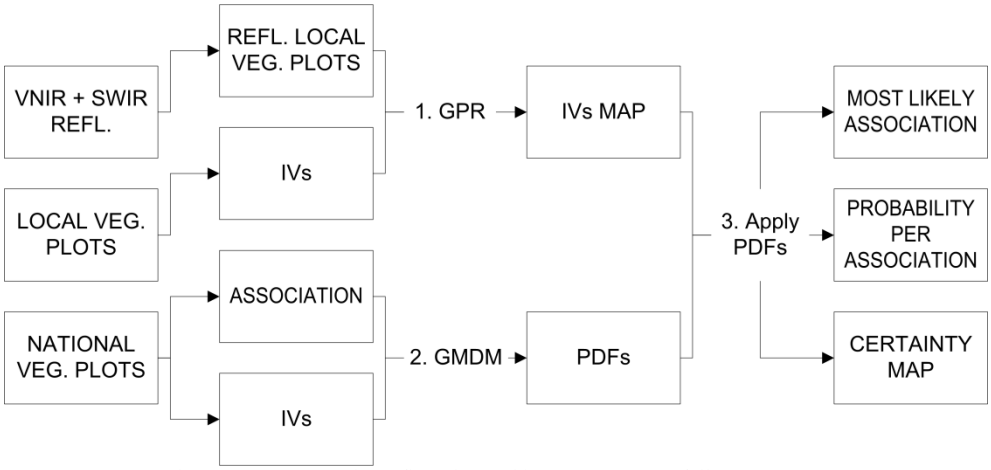

Figure 5.1 Research structure presented in a flow chart. Abbreviations are as follows: RS = remote sensing, veg = vegetation, as in vegetation plots and vegetation maps, Refl = 3.5 m spatial resolution VNIR and SWIR reflectance data from airborne optical AHS sensor, IVs = indicator values for salinity (mS), moisture (mF) and nutrients (mN), GPR = Gaussian Process Regression, GMDM = Gaussian Mixture Density Modelling and PDFs = Probability Density Functions, See section '5.3.1. General Methodology' for further information.

5.3.2. Study site

The island of Ameland, located in the Dutch part of the Wadden Sea, was selected as study area. The Wadden Sea is located southeast of the North Sea, stretching from the Netherlands to Denmark; it is an intertidal zone of great ecological importance, as is reflected by its assignment as a Natura 2000 area and as an UNESCO world heritage site. Ameland (60 km^2 , 53.45°N, 5.684°E, Fig. 5.2) is the third largest Dutch island in the Wadden Sea and has a large variety of ecosystems, ranging from dry and wet dunes, to tidal salt marshes, heath lands

and fresh water marshes. Natural gas has been produced from an extraction facility (Fig. 5.2) since 1986, causing soil subsidence (Dobben & Slim, 2012). The high number of abiotic gradients that was expected to be present and the wide variety of ecosystems on the island were the main motivation for selecting Ameland as a study site. Additional reasons were that sufficient local vegetation plots were available and that nature management practices have been stable over the last decades (at least since 1960). Only small changes in vegetation have been observed since 1986; some loss of diversity and a shift towards wetter species, mainly caused by succession, weather fluctuations and soil subsidence due to gas extraction and eutrophication (Dobben & Slim, 2012). In this study we focused on a study site of 12 km^2 located on the eastern end of the island, containing mainly short natural vegetation dominated by graminoids, herbs, dwarf shrubs, mosses and lichens. Some prominent features of the study site are highlighted in Fig. 5.2.

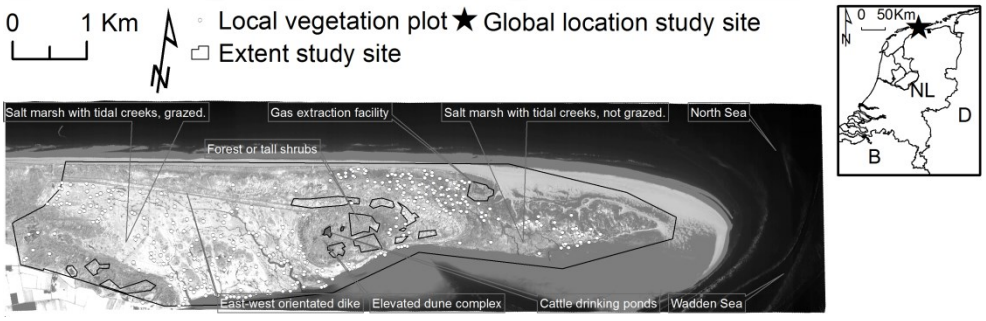

Figure 5.2 Coverage of the AHS remote sensing image and delineation of the study site, including the global location of Ameland. Shown here is the eastern extent of the flight strip, which completely covers the study site delineated with the black line. Areas with forests and high shrubs were manually identified and removed. Water, bare soil, sand and artificial features were masked within this area, resulting in the final study site that appears in Figs. 4 and 5. Prominent features are indicated in the Fig. and include the two salt marshes with tidal creeks, the elevated dune complexes and the beach extending along the north and east edges of the island. The locations of the 379 local vegetation plots are indicated with white circles.

5.3.3. Acquisition and preparation of vegetation data

For step1 (Fig. 5.1) we used 379 vegetation plots located in or near the study site (hereafter: 'local plots') for which the floristic composition, plot location and dimensions were known (Table 5.1). These local plots originated from two existing sources (Jager, 2010; Slim et al., 2005), supplemented with afield campaign in 2010 to gather a sufficiently large number of vegetation plots. The majority of the plots (86%) were surveyed by more than one field ecologist to reduce observer bias (i.e. systematic misidentification of species (Rocchini et al., 2012)). Species records for all plots were corrected for synonyms of plant species names. Some database features are different in the three data sources, such as the sampling strategy (Table 5.1). Also, the area of investigation was different for each source, but the extent of the 2010 field campaign encompassed the areas of the two other sources. Plot surveys were carried out in different years, but only small changes in the vegetation have occurred during recent years (Dobben & Slim, 2012). In all, we felt confident to use plot data from different

sources because only species composition, plot dimension and location were used and these data were consistently recorded for each plot.

For step 2 (Fig. 5.1) we had access to the Dutch National Vegetation Database (GIVD ID: EU-NL-001, (Schaminée et al., 2012). From this we extracted a subset of 35 000 plots (hereafter: 'national plots') that was earlier classified into an association during compilation of the standard work of plant associations in the Netherlands (Schaminée et al., 1996; Schaminée, Stortelder, et al., 1995; Schaminée, Weeda, et al., 1995a, 1995b; Stortelder et al., 1999). The national plots are distributed across the whole of the Netherlands. There was no overlap between the national and local plots.

Considering the strong influence of tidal movements and elevation differences between the mudflats and dune tops, we expected a large range of variation in salinity, moisture regime and nutrient availability of the soil. We therefore chose to use IVs for soil salinity (mS, ranging from 1 = fresh to 3 = saline), moisture regime (mF, ranging from 1 = aquatic to 4 = dry) and nutrient regime (mN, ranging from 1 = nutrient poor to 3 = very nutrient rich) as explanatory variables for association occurrence. From a previously existing list of IVs per plant species (Witte et al., 2007) we gathered the values of mS, mF and mN for each plant species occurring in the national and local plots. A plot averaged value for mS, mF and mN was calculated as the arithmetic mean of the IVs of all species. Abundance of a species within a plot was not used as weighing factor for IV averaging, as this does not lead to an increase in the extent to which an IV is representative for actual site conditions (Käfer & Witte, 2004).

Table 5.1 Origin and number of the local vegetation plots.

Source	n plots	Surveyed	Coordinates accuracy	Sampling strategy	Plot area	n surveyors
Field Campaign	53	Aug- Sept 2010	< 1 cm (RTK-DGPS)	Along nutrient, salinity and moisture gradients	$4m^2$ ($25m^2$ for dwarf shrub plots, n = 5)	1
(Jager, 2010)	213	July, Aug. Sept. 2009	< 5 m	Systematic following vegetation units delineated on aerial photo	$3m^2$ to $25m^2$	6, working mainly in pairs
(Slim et al., 2005)	113	Aug. 2004	< 1 cm (RTK-DGPS)	Random	$4m^2$	2, always paired
Total	379					

5.3.4. Acquisition and preparation of remote sensing data

An Airborne Hyperspectral Scanner (AHS) image, dated June 19 2005 0900h GMT+1 covered the study site in a single flight line from east to west. No clouds were present in the image. Geometric corrections were applied according to the standard procedures of VITO (Biesemans et al., 2010), using the SRTM DEM. Atmospheric correction was performed

using the MODTRAN4 radiative transfer code (Biesemans et al., 2007; De Haan et al., 1991), resulting in top of canopy reflectance values.

The image consists of spectral information in 21 bands. The first 20 bands cover the visible and near infrared (VNIR, 430–1030 nm) regions with a full width half maximum (FWHM) of 28 nm. Band 21 covers part of the shortwave infrared (SWIR, 1500–165 0 nm) region with 150 nm FWHM. Originally, 42 additional SWIR bands were provided, but these were discarded due to a low signal to noise (SN) ratio. Ground resolution was 3.5 m (Gómez et al., 2007).

Given that none of the local plots included forests or tall shrub, pixels with such vegetation structures were visually identified using topographic maps, detailed aerial photographs and field observations and excluded from the study site. Clearly identifiable objects, such as the gas extraction facility and two cattle drinking ponds, were also manually removed. Pixels with NDVI < 0.35 were assumed to contain bare soil, water, paved areas or a mixture of these and were removed (a value of 0.35 was high enough to remove bare dunes, sea, tidal creeks and some of the paved areas, but still low enough to retain the vegetated areas). We imposed a circle with 2.5 m radius around the centre of each plot to account for geometric inaccuracies of the pixels and the GPS recorded coordinates. Plot averaged reflectance values were generated as the mean reflectance of all pixels that intersected this circle. On average, the vegetation plots intersected with 1.5 pixels so the contribution per pixel to the plot average was not weighted

5.3.5. Predicting IVs with Gaussian process regression

IVs of a random 80% of the local plots were related to the local plot reflectance values using Gaussian Process Regression (GPR) (step 1, Fig. 5.1), a nonparametric and nonlinear machine learning regression technique (Pasolli et al., 2010; Rasmussen & Williams, 2006). GPR formulates the learning of the regressor within a Bayesian framework, where the regression model is derived by assuming that the model variables follow a Gaussian prior distribution encoding the prior knowledge about the output function (Pasolli et al., 2010). Key features of GPR include the relatively small training set required while developing a regression model, the flexible kernels, identification of relevant bands during the calibration and generation of pixel wise confidence intervals (Verrelst, Alonso, et al., 2012). A major motivation to work with GPR instead of the more commonly used Partial Least Squares Regression (PLSR, (Wold et al., 2001)) is that earlier experiments evaluated GPR as a very competitive method when retrieving biophysical parameters such as leaf chlorophyll content, leaf area index or fractional vegetation cover, compared to vegetation indices (Verrelst, Alonso, et al., 2012) or other machine learning algorithms such as Neural Networks, Support Vector Regression and Kernel Ridge Regression (Verrelst, Muñoz, et al., 2012). Furthermore, GPR is more capable of dealing with the expected collinearity of the predictor variables, where the nonlinearity of the technique allows for saturation of the relations near the minimum and maximum values of the IVS. For comparison, PLSR models were also fitted to predict the IVs, where the number of latent variables was chosen as to

minimise the Root Mean Square Error (RMSE) of the Leave One Out (LOO) validation (Schmidtlein, 2005).

A separate GPR model was fitted for each IV, using the GPML toolbox in MatLab (version 3.2; (Rasmussen & Williams, 2006)). The GPR models were validated by predicting IVs for the remaining 20% of the local plots, yielding a coefficient of determination R^2 and RMSE. The developed GPR models were applied pixel wise to the study site to generate IV prediction maps. In addition the relevance of each spectral band and the prediction standard deviation per pixel as an indication of the spatial variation in prediction confidence were recorded.

5.3.6. Modelling association occurrence probability

We selected19 associations for modelling in the study site (Table 5.2). The selection aimed to represent the dominant vegetation and floristic diversity in the study site and was based on available literature (Dobben & Slim, 2012; Westhoff & van Oosten, 1991), observed associations in local vegetation plots (at least 48 different associations were observed in the local plots) and advice from experienced ecologists familiar with the Dutch phytosociological system and the study site. We were unable to include all associations that were known to be present, because then unique niches in IV-space would probably be occupied by more than one association. We therefore focussed on the most prevalent associations and left sparsely occurring associations out of consideration. Also, from groups of very floristically similar associations, one representative association was selected (for example, 25AA01 *Salicornietum dolichostachyae* and 25AA02 *Salicornietum brachystachyae* differ only with regard to the dominant Salicornia species: *Salicornia europaea* or *Salicornia procumbens*). The selection was further constrained to associations that were sufficiently represented in the national database to allow for reliable calculation of the Probability Density Functions (PDFs) and to associations that do not contain tall shrubs or trees. Finally, so called 'lump associations', which are defined by virtue of lacking certain diagnostic species, were not selected because such associations are considered of less floristic relevance.

We thus selected those vegetation plots from the national database for which the association matched one of the 19 target associations (3944 plots) and extracted the averaged IVs and association. The rationale behind our modelling approach is that vegetation plots of a given association have a more or less similar floristic composition, as dictated by the syntaxonomy of the DVN. As the plot-averaged IV was determined from individual species IVs, such plots will have comparable mean IVs, causing associations to appear as clusters of vegetation plots in a 3-dimensional IV space.

For each of the selected associations, we fitted PDFs through the IV-space. Note that this was done with the national plots, thus independently from the local plots and remote sensing data (step 2, Fig. 5.1). This is a major advantage over conventional classification algorithms that require training areas of the desired vegetation units. We used Gaussian Mixture Models (GMMs) to model the PDFs. GMMs can approximate any shape, which is a valuable attribute here given the non-symmetric shapes of many associations in IV-space. The PDFs were thus fitted independent of a presumed shape so that the correlation among

IVs is accounted for. Based on the PDFs thus obtained, we were able to calculate a Bayesian occurrence probability of each association as a function of IVs. The accuracy of the calibration was assessed by fitting PDFs again, but now with one randomly selected half of the 3 944 national vegetation plots and calculating association occurrence probability for the remaining plots of the national dataset. Comparing the observed association with the most likely occurring association yielded a model efficiency E (percentage of correctly classified plots) and the chance corrected kappa coefficient k (Cohen, 1960). One association (12BA04 *Ononido-Caricetum distansis*) was not involved in this procedure, as the number of plots of this association in the selected national plots did not meet the minimum number recommended to accurately fit and validate a PDF (30 plots, (Douma, Witte, et al., 2012)). The PDFs were fitted using the PARDENS, v1.10b software (Wójcik & Torfs, 2003).

Table 5.2 Selected associations for predictions, taken from the Dutch national phytosociological system. The code consists of a sequence of indicators for the hierarchical syntaxa of thephytosociological system. A diagnostic species for each association is provided. A detailed description and a complete list of character and differential plant species can be found in Schaminée, Sýkora, Smits, and Horsthuis (2010).

Code	Association name	Description	Diagnostic species
06AC04	*Samolo-Littorelletum*	Rare, pioneer, association with short plants, found in young dune valleys with high pH and high groundwater levels.	*Samolus valerandi*
08BB04	*Typho-Phragmitetum*	Very common, species poor association consisting of 2-4 m high reed beds. Found in fresh to brackish water.	*Phragmites australis*
12BA02	*Triglochino-Agrostietum stoloniferae*	Species rich, grassland vegetation, found on peaty soils, bordering moist dune valleys. Very common.	*Triglochin palustris*
12BA03	*Trifolio fragiferi-Agrostietum stoloniferae*	Low grassland vegetation, found in brackish environments just above floodplains or in dune valleys. Often grazed.	*Trifolium fragiferum*
12BA04	*Ononido-Caricetum distantis*	Association consisting of small, thorny shrubs. Rare. Found at sporadically inundated locations with high pH.	*Ononis repens*
14AA02	*Violo-Corynephoretum*	Low pioneer vegetation with many mosses and lichens. Found in nutrient poor, dry and slightly acid sandy soils, such as south facing dune slopes.	*Cladina ciliata*
14BB02	*Festuco-Galietum veri*	Common association of dune	*Hypochaeris radicata*

		grasslands with mosses, found in a dry and nutrient poor environment.	(diagnostic for order level. This association lacks species diagnostic for the association level).
14CA01	*Phleo-Tortuletum ruraliformis*	A common, very open pioneer association, characterised by mosses and lichens as dominant species found in calcium rich dunes.	*Erodium lebelii*
20AB04	*Pyrolo-Salicetum*	Association consisting of small shrubs, found in dry to moist dune valleys. Species rich.	*Pyrola rotundifolia*
23AB01	*Elymo-Ammophiletum*	Association found on high sand dunes, characterised by bare sand and high grass tussock that capture sand-drift and forms new dunes.	*Ammophila arenaria*
24AA02	*Spartinetum townsendii*	Species poor association in wet and alkaline conditions. Found at the lower parts of salt marshes and along tidal creeks. Consists of dense grass tussocks of up to 1 m. high.	*Spartina anglica*
25AA01	*Salicornietum dolichostachyae*	Very species poor, low vegetation found on lower parts of salt marshes along the high-water mark.	*Salicornia procumbens*
26AA01	*Puccinellietum maritimae*	Common association, species poor dense grass mats that float when inundated. Found on low salt marshes. Requires cattle grazing to sustain.	*Puccinellia maritima*
26AA02	*Plantagini-Limonietum*	Species poor and rare association that is often inundated and found on low and central parts of the salt marshes.	*Limonium vulgare*
26AC01	*Juncetum gerardi*	A common association consisting of low and highly compact vegetation, found on more elevated and occasionally inundated parts of the salt marshes.	*Juncus gerardii*
26AC03	*Junco-Caricetum extensae*	Relatively species rich, open and moderately high	*Carex extensa*

		association, found at the foot of the dune strip, just above the high water mark.	
26AC05	*Artemisietum maritimae*	Species poor association, consisting of dense 40-50 cm high silver-grey coloured plants, found on local elevations in the salt marshes.	*Seriphidium maritimum*
27AA01	*Sagino maritimae-Cochlearietum danicae*	Rare association, consisting of very low and open pioneer plants.	*Sagina maritime*
27AA02	*Centaurio-Saginetum*	Pioneer association in dune valleys. Species rich, very low plants and open structure.	*Centaurium littorale*

5.3.7. Remotely sensed vegetation and occurrence probability maps

The GPR predicted pixel IVs were combined with the PDFs of each association to calculate the occurrence probability of each association for the complete study site (step3, Fig. 5.1). This yielded occurrence probability maps per association on a scale from 0 to 1. After assigning each pixel to the association with the highest occurrence probability, a crisp map of predicted associations per pixel was obtained. In addition, the spatial prediction certainty was mapped using the maximum occurrence probability per pixel.

Several alternatives were considered to assess the quality of the resulting predictions of association occurrence. A standard classification algorithm such as nearest neighbour classification was not suitable because insufficient training pixels could be identified in the image. The local plots were not suited for validation purposes either, because only approximately 37% of the local plots were unequivocally classified to an association (data not shown) and not all modelled associations were found in the local plots. Moreover, the local plots did not amount to the recommended minimum number of plots per association required for a reliable validation (Douma, Witte, et al., 2012). This is explained by the fact that our method produces occurrence probabilities of several associations for a given set of IVs, not binary presence or absence. This contrasts to a hypothetical set of validation vegetation plots which consist of presence/absence of associations in a given plot. Therefore one plot should be considered as one of the possible realisations determined by the multinomial probability distribution of the model. Assuming the model is correct, the occurrence frequency will not converge to the probability of occurrence until the number of plots is sufficiently large. Douma, Witte, et al. (2012) describe how, with increasing number of validation plots, the occurrence frequency of each association approaches the occurrence probability. Therefore, a certain minimum number of plots (around 30) per association is required to appreciate the extent to which the predicted occurrence probabilities match the actual occurrence probability.

Instead, a local vegetation expert with detailed floristic knowledge of the study site was asked to assess the extent to which the modelled patterns, for both the crisp vegetation map with the most likely occurring association per pixel, as well as the occurrence maps per association, matched reality.

In addition, the predictions of occurrence probability per association were compared to a preexisting vegetation map ((Jager, 2010), hereafter: the preexisting vegetation map). This preexisting vegetation map was commissioned by the Dutch Ministry of Public Works as part of the continuous monitoring of coastal areas. The map was produced following the 'photoguided approach' (Küchler & Zonneveld, 1988), meaning that homogeneous patches of vegetation were delineated on high resolution false colour aerial photographs. Subsequently, the delineated areas were interpreted and labelled using vegetation plots and field observations. The vegetation units of the preexisting vegetation map were translated to the associations used here based on a key provided in the map's documentation (Jager, 2010).

Comparison of our model results with the preexisting vegetation made was only visual and qualitative for a number of reasons. Firstly, it appeared that only 12 of the 19 modelled associations were present in the map due to the fact that the preexisting vegetation map's extent is smaller than the study site. Secondly, the preexisting vegetation map does not present distribution of individual vegetation units. Instead, the polygons of the preexisting vegetation map were labelled with legend units that were specific compositions of multiple vegetation units. Creating a crisp map of associations was therefore not feasible and we computed the share of each association within each map polygon on an ordinal scale. The different measuring scales prevented systematic comparison of the preexisting vegetation map (ordinal) with the final map (nominal scale, Fig. 5.5) or with the occurrence probabilities per association maps (continuous scale, supplementary data in section 5.8.3). Finally, a high predicted occurrence probability does not indicate actual occurrence per se. The location can be occupied by an association even more likely to occur, or by a less suitable association due to chance events. Only for relatively large clusters of pixels with consistent high occurrence probability can probability be expected to converge to occurrence frequency.

5.4. Results

5.4.1. Estimating IVs with Gaussian process regression

Nearly the full range of IVs was observed in the local vegetation plots (Fig. 5.3). Only for mF was the theoretical minimum (mF = 1) not obtained, meaning that no open water was present in the vegetation plots. Abiotic extremes of nutrient poor to nutrient rich (mN1 to mN 3), highly saline to fresh water (mS 1 to mS 3) and wet to dry conditions (mF 1.2 to mF 4) were all present in the local plots.

Validation of the GPR models yielded R^2 values 0.73, 0.64 and 0.76 and RMSE of 0.31, 0.38 and 0.35 for mS, mF and mN respectively (Fig. 5.3). Despite correlations among the IVs,

selected relevant bands strongly differed among the IVs (supplementary data in section 5.8.1), suggesting unique spectral signatures for each IV. Predicted values were generally on par with the observed values although considerable variation around the 1:1 line was observed. All three IVs, but especially mF, appeared to be under predicted for higher values and over predicted for lower values (Fig. 5.3). The observed range of IVs was covered by the GPR-predicted IVs, but for mS and mN, the range of predicted IVs slightly exceeded the observed range. Model residuals of both the training and validation data were distributed normally around zero (results not shown).

Applying the developed GPR models to the AHS data yielded full coverage IV maps (Fig. 5.4). For approximately 88% of the pixels in the study site, all three IV predictions were within the limits of the IV scale (1–3 for mS and mN, 1–4 for mF). The predicted IVs resembled a logical spatial arrangement of variation in salinity, nutrient availability and moisture regime. The sea influence in the salt marshes is apparent from the low mF and high mS values, which gradually increased (mF) and decreased (mS) with distance from the sea and the tidal creeks. A more abrupt transition is apparent where the saltmarshes are bordered by a dike or where the dunes rise steeply. The elevated dune complexes themselves were drier (high mF) and less nutrient rich (low mN). While the surroundings of the cattle drinking ponds in the centre of the central dune complex were correctly assigned to low mF (i.e. moist soil) values, mS was over predicted for those areas (i.e. too high salinity). Low mF values north of the central dike match expectations, as these areas are subject to in frequent flooding. High standard deviations of the GPR predictions are found along transition from sea to salt marsh, along the tidal creeks and around the cattle drinking ponds, as well as in several small patches in the western part which are likely small ponds or wet areas (supplementary data in section 5.8.2). These areas are likely to possess spectral characteristics that were insufficiently represented in the training data. For these areas, uncertainties were found to be high for all IV models (Fig. 5.8).

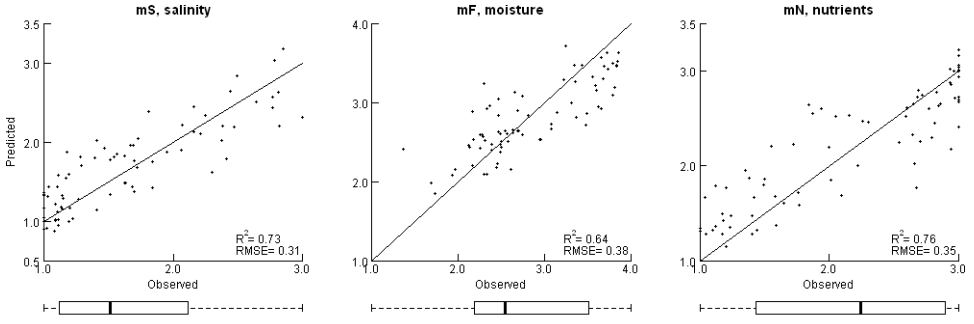

Figure 5.3 Validation of the GPR IV predictions, using validation set of 20% of the local plots. Variation of observed IVs in the local plots is expressed in the horizontal box plots, where the whiskers indicate upper and lower quartiles and outliers are given as circles. Note that the 'predicted' axes extend beyond the observed range for mS and mN because the predicted values exceed the minimum and maximum observations.

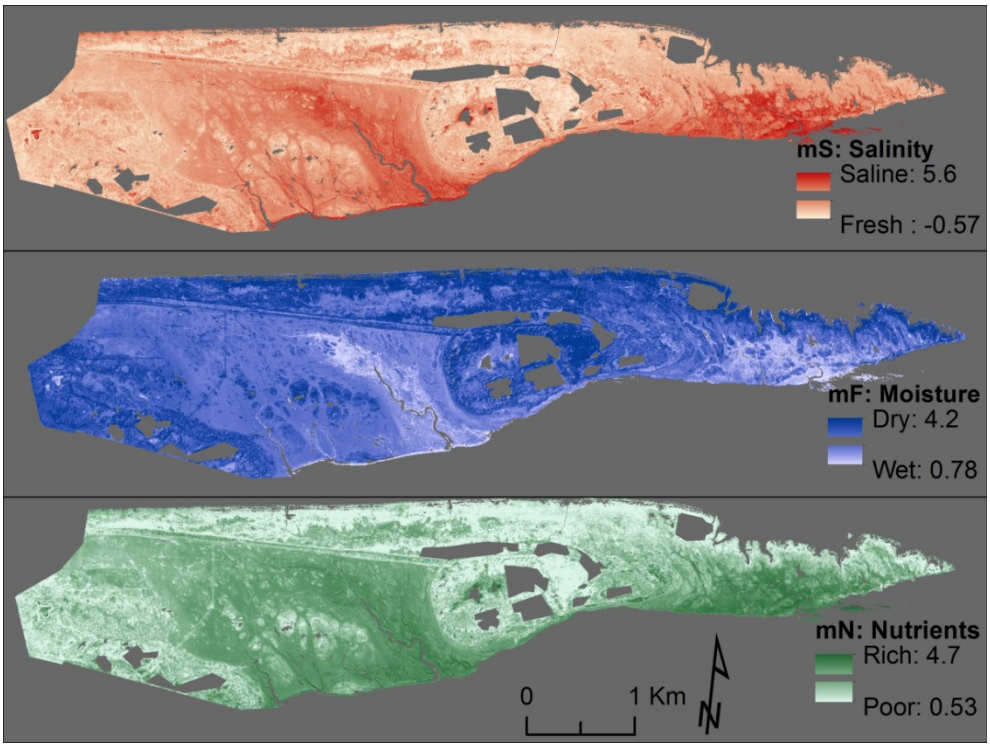

Figure 5.4 GPR models applied to the study site, showing IVs for salinity, moisture and nutrient regime. The saline, nutrient-rich and frequently flooded salt marshes are easily distinguished from the drier and less saline dunes. Sea influence is most prominent along the southern shore, where the sea is not blocked by high dunes as it is along the north coast. IV estimations extend beyond the theoretical limits (1–3 for mS and mN, 1–4 for mF) for 6, 0.01 and 8% of the study site for mS, mF and mN respectively. For approximately 88% of the study site, all three IV estimations were within the theoretical bounds

5.4.2. Modelling association occurrence probability

Based on PDFs of association occurrence as a function of IVs, associations were reasonably well discriminated: 60% of the validation plots were correctly classified (kappa = 0.46, Table 5.3). Despite the overall reasonable discrimination, associations 14AA02 *Violo-Corynephoretum*, 14BB02 *Festuco-Galietum very* and 14CA01 *Phleo-Tortuletum ruraliformis* were often confused with one another, while also often being incorrectly considered as 20AB04 *Pyrolo-Salicetum* or 23AB01 *Elymo-Ammophiletum*. 24AA02 *Spartinetum townsendii* was often predicted but hardly ever correctly. Instead, 25AA01 and the associations from class 26 were often mistaken for 24AA02. Association 26AA02 *Plantagini-Limonietum* was predicted as 26AA01 *Puccinellietum maritimae* three times as often as being predicted correctly.

5.4.3. Vegetation maps and occurrence probability maps

After combining the PDFs of the associations with the GPR-predicted IVs, occurrence probabilities per association were assigned to approximately 99% of the study site pixels. No probabilities were assigned to a pixel when the summed occurrence probabilities at that specific IV combination were lower than 0.0001. In nearly all cases, this coincided with pixels where at least one of the three IV predictions exceeded the theoretical range of IV or that had a high predictive standard deviation in the GPR modelling.

Comparison of the occurrence probability per plant association to occurrence according to the preexisting vegetation map suggested mismatches for several associations ((Jager, 2010), supplementary data in section 5.8.3). For 12BA04, 27AA01 *Sagino maritimae-Cochlearietum danicae* and 27AA02 *Centaurio-Saginetum*, high occurrence probabilities were noted throughout most of the study site even though observations were limited to just a few areas. Other associations had high probability occurrences in parts of the map where they did not occur in reality (12BA03 *Trifolio fragiferi-Agrostietum stoloniferae* and 26AA01 at the eastern end of the study site, 26AC01 *Juncetum gerardi* and 26AC05 *Artemisietum maritimae* at the central salt marsh). Occurrence of 24AA02 was correctly estimated around the southern edge of the central salt marsh, but high occurrence probability around the eastern tidal creek was not supported by observations. Some associations had low or zero occurrence probabilities, even though they were known to be present (08BB04 *Typho-Phragmitetumin* the northwestern part of the central salt marsh, 26AA02 under predicted at the eastern salt marsh). Nonetheless, some areas are correctly predicted to likely host a certain association (26AA01 at the central salt marsh, 26AC03 just north of the central east–west orientated dike and 26AC05 along the tidal creek in the central salt marsh and on the eastern salt marsh).

Table 5.3 Confusion matrix of the validation of the Probability Density Function (PDF) fitting, where a random half of the national plots were used to calibrate the PDFs and subsequently used to calculate the most likely association given the combination of IVs for each plot in the remaining half of the national plots. The observed associations for the validation plots are indicated in rows, while the predicted associations are in the columns. User's accuracy (U.A., number of correctly classified plots for association y / total number of plots predicted as y) and producer's accuracy (P.A., number of correctly classified plots for association y / total number of plots observed as y) are given per association.

		Predicted																			
Observed		06AC04	08BB04	12BA02	12BA03	14AA02	14BB02	14CA01	20AB04	23AB01	24AA02	25AA01	26AA01	26AA02	26AC01	26AC03	26AC05	27AA01	27AA02	Total	Correct	P.A.
06AC04	*Samolo-Littorelletum*	40	2	1					1											44	40	0.91
08BB04	*Typho-Phragmitetum*	3	134	9															1	147	134	0.91
12BA02	*Triglochino-Agrostietum stoloniferae*		1	28	4														1	34	28	0.82
12BA03	*Trifolio fragiferi-Agrostietum stoloniferae*		1	1	54										5	3		2	3	69	54	0.78
14AA02	*Violo-Corynephoretum*					129	15	52	3	3										202	129	0.64
14BB02	*Festuco-Galietum veri*					7	91	10	3	3									1	115	91	0.79
14CA01	*Phleo-Tortuletum ruraliformis*					50	14	146	1	1										212	146	0.69
20AB04	*Pyrolo-Salicetum*						2		24										2	28	24	0.86
23AB01	*Elymo-Ammophiletum*				1	4	3		1	72								5		86	72	0.84
24AA02	*Spartinetum townsendii*										41	14	7	7	1					70	41	0.59
25AA01	*Salicornietum dolichostachyae*										17	15	3							35	15	0.43
26AA01	*Puccinellietum maritimae*				11						74	6	79	188	41	5	28	1		433	79	0.18
26AA02	*Plantagini-Limonietum*										5		6	63			3			77	63	0.82
26AC01	*Juncetum gerardi*				9						10		33	4	63	15	15	2	2	153	63	0.41
26AC03	*Junco-Caricetum extensae*				2						1		2		7	27	2	1	4	46	27	0.59
26AC05	*Artemisietum maritimae*												1	5	1		31	2		40	31	0.78
27AA01	*Sagino maritimae-Cochlearietum danicae*				2		2			4					3	1	4	70	6	92	70	0.76
27AA02	*Centaurio-Saginetum*			1	5				1						1			10	60	78	60	0.77
	Total	43	138	40	88	190	127	208	34	83	148	35	131	267	122	51	83	93	80			
	Correct	40	134	28	54	129	91	146	24	72	41	15	79	63	63	27	31	70	60			
	U.A.	0.93	0.97	0.70	0.61	0.68	0.72	0.70	0.71	0.87	0.28	0.43	0.60	0.24	0.52	0.53	0.37	0.75	0.75			

A crisp map of associations, based on the most likely occurring association for each pixel, showed widely varying distributions of associations (Fig. 5.5). Only six associations occupied each more than 5% of the map, with one association (12BA04) accounting for nearly 25%. The remaining 13 associations occurred rarely or not at all (06AC04 *Samolo-Littorelletum*) according to the model calculations (Fig. 5.6). The associated probability of the most likely association shows the certainty of our predictions. Low maximum probability values indicate areas where a mixture of several associations is likely to occur (e.g. in the central saltmarsh), whereas high values indicate that a single association is by far the most likely (for example as occurring in dune complexes) Certainty of the predictions is homogeneous throughout the dune complexes, while the salt marshes have slightly lower certainty. Areas with high prediction standard deviations do not necessarily have low maximum probability values. Uncertainties of the GPR models are not accounted for in the predictions of occurrence probabilities of associations, which may explain the lack of coincidence.

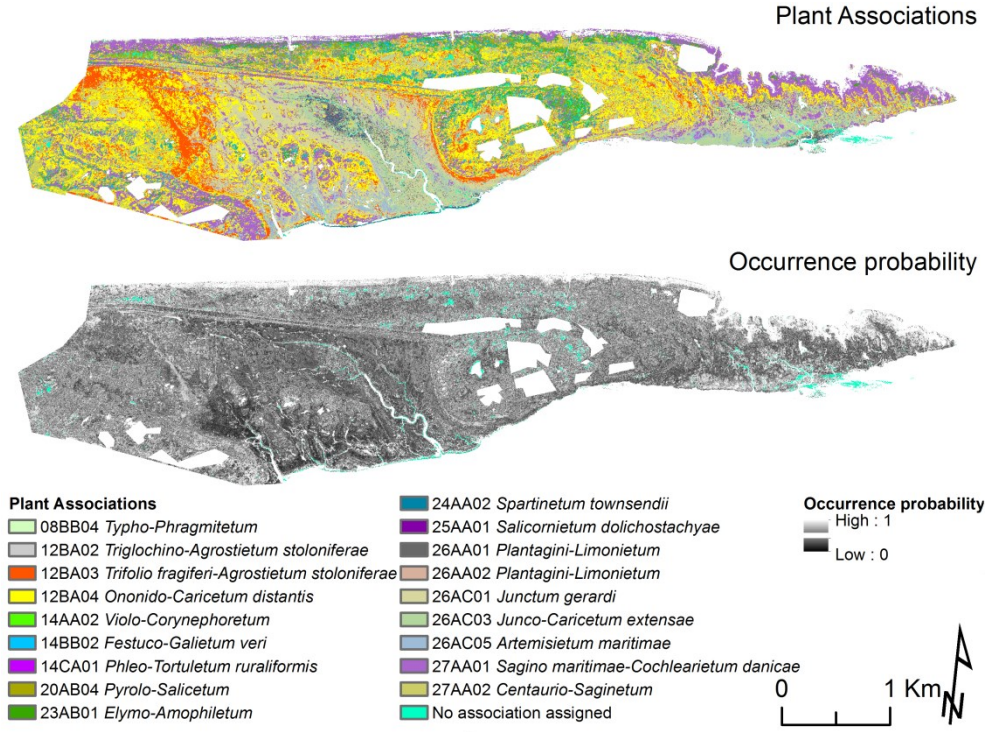

Figure 5.5 Most likely occurring association visualised per pixel (top) and the associated value of the maximum occurrence probability (bottom).

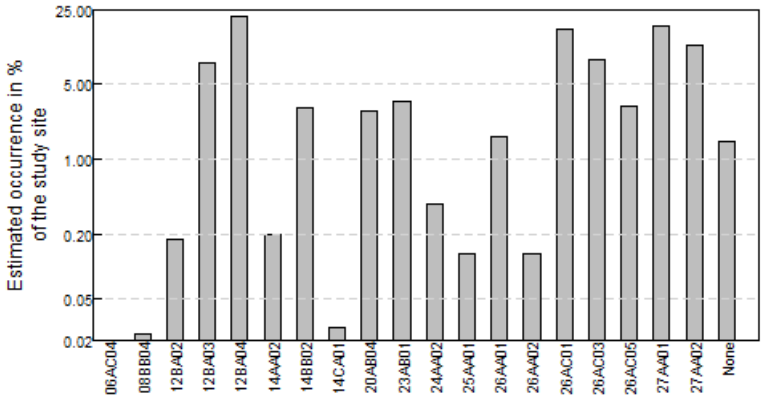

Figure 5.6 Barplot of the area of the predicted associations in % of the study site. Note log scale on the y-axis.

5.5. Discussion

5.5.1. Advantages of our approach

In this study, we evaluated a method that uses VNIR and SWIR airborne reflectance data to predict the spatial extent of associations that have been defined a priori and formulated based on species composition. Central to our approach is the inclusion of vegetation characteristics, in this case IVs, as an intermediate between remote sensing data and association occurrence in order to allow predictions. This approach builds on two key techniques from earlier studies: IV predictions from reflectance data were demonstrated in, among others, Schmidtlein (2005) while Witte et al. (2007) pioneered association predictions based on IVs. This study is the first attempt to merge these two techniques and aimed to further advance a priori defined associations into the remote sensing domain. In addition to association occurrence, our method produces maps of IVs for environmental factors to which vegetation responds, in this case salinity, moisture regime and nutrient availability. These maps present useful information on abiotic conditions and its impacts on vegetation distribution to local nature managers and policy makers (Schmidtlein, 2005).

While the prediction of IVs is based on local vegetation plots, the PDFs of the associations were calibrated using an independent database of vegetation plots (the national plots). Hence, the calibration is independent of the study site, local plots and remote sensing source, which is a major advantage, compared to common classification techniques. This also ensures that the method can be applied to other areas. Because the calibration data remains unchanged, the predicted vegetation relations are generic and consistent between areas and studies. Once the IVs are predicted, the user is free to select any set of relevant vegetation units from the national database. This can be units from different hierarchical levels of the phytosociological system (associations, alliances, orders or classes) or from any other classification of vegetation units. The only condition is that the classification is based on the floristic composition of plots, as recorded in the national database.

Another advantage is that the method does not force the user to present the occurrence of vegetation units as a binary present/absent state, but gives occurrence probabilities on a continuous scale. This does justice to actual spatial patterns in plant cover, which are usually gradual rather than abrupt due to gradients in environmental factors (such as the availability of moisture and nutrients), history and stochastic processes such as colonisation, disturbance and dispersal (Fukami & Wardle, 2005; Maarel & Sykes, 1993; Ozinga et al., 2005). In addition, the maximum occurrence probability for a pixel over all the associations reveals if one association is clearly the most likely to occur (high certainty), or if several associations are likely and one is only marginally more likely than the remaining (low certainty).

5.5.2. Critical evaluation of results – prediction of IVs

We predicted IVs for salinity (mS), moisture regime (mF) and nutrient availability (mN) using Gaussian Process Regression. While there is no physical basis for a relationship between IVs and spectral reflectance, we anticipated a relation because IVs reflect both the environmental conditions of a site as well as plant traits selected by those conditions and both are represented in spectral reflectance. Indeed, the correlation between the spectra and the plot IVs suggests that spectral reflectance changes consistently between plots with different IV combinations. Thus, IVs may be considered an umbrella concept that summarises relevant vegetation properties that directly influence canopy reflectance and that reflect vegetation functioning.

The accuracy of the IV predictions (RMSE = 0.31, 0.38 and 0.35, respectively, for mS, mF and mN) in this study is within the same order of magnitude as in earlier studies. Ecker, Waser, and Küchler (2010) predicted mF and mN from different spectral and LiDAR sources, achieving RMSE ranging from 0.10 to 0.32 and 0.12 to 0.40 for mF and mN respectively. Schmidtlein (2005) predicted mF and mN from hyperspectral VNIR imagery with RMSE of 0.89 and 1.13, respectively, while Schmidtlein and Sassin (2004) predicted mF, mN and mR with RMSE of 0.43, 0.24 and 0.31, respectively. To our knowledge, this study presents the first prediction of mS from remote sensing data, as well as the first instance of using GPR for IV prediction. GPR acts as a very competitive method and indeed, IV predictions based on Partial Least Squares Regression (PLSR) in our study site were clearly less accurate (R^2 = 0.68, 0.58, and 0.68 and RMSE = 0.37, 0.41, and 0.80 for mS, mF and mN, respectively) compared to the GPR approach, suggesting that, at least in this case, GPR was more suited to predict IVs. Our predictions were obtained using hyperspectral data with 20 spectral bands in VNIR and one in the SWIR region. On average, about half of the bands contributed to the observed IVs (supplementary data in section 5.8.1), suggesting that similar predictive power may be obtained with fewer bands, but only if those bands are known to encompass most of the variation. On the other hand, having more bands available does not negatively impact the regression model; less important bands are down weighted in the regression model (Rasmussen & Williams, 2006). Hence, it does no harm having 21 bands available in the regression model. By down weighting non relevant bands, the GPR algorithm effectively also reduces collinearity.

The single SWIR band was consistently important for all three IV models (supplementary data in section 5.8.1), whereas the other influential bands were different for each model suggesting IV-specific selection of spectral bands. Little agreement is observed between the spectral regions relevant for IV prediction in this study and those in Schmidtlein (2005), all the more since SWIR spectral information is absent in the latter.

The weakest performing IV appeared to be with moisture availability. This may be due to contrasting spectral information for similar mF values (Schmidtlein, 2005). Low mF values are found in both nutrient rich, saline locations such as the salt marshes and in nutrient poor and non-saline locations such as dune valleys. Low mF values are thus related to different spectral responses that may not be separated by GPR. For other IVs, generally consistent IVs to reflectance relationships seem to be at work. For example, in this study area, high mS value were consistently related to proximity to the sea, which in turn translates to remotely sensible properties such as decreasing vegetation density, soil substrate and soil moisture (Mulder et al., 2011). Likewise, mN is related to biomass, productivity and leaf area index (Diekmann, 2002; Ertsen et al., 1998; Klaus et al., 2012; Schmidtlein, 2005), which are known to be detectable by spectral observations (Asner & Martin, 2008; Tucker, 1979). In this study site, biomass is relatively low as forests and high shrubs were omitted. Instead, high nutrient availability is always linked to sedimentation processes. While the correlations among the IVs were strong (Pearson's r = -0.83, 0.86 and -0.78 for mS–mF, mS–mN and mF–mN, respectively), we do not think this greatly affected the IV maps. The differences in band selection (supplementary data in section5.8. 1) between the various IVs suggest that confounding influences are relatively small. This is also apparent from the large differences in the respective IV maps. Instead, the uncertainties in IV predictions may be attributed to changes in vegetation density and composition, effects of flooding events during the time lag between vegetation survey and image acquisition and to interfering effects of vegetation structure, background and shadow and of course to the reliability of the IVs themselves.

5.5.3. Critical evaluation of results – modelling association occurrence probability

Following IV predictions, the subsequent part of our approach consisted of modelling probability density functions of association occurrence as a function of IVs. For this, we hypothesised that each association occupied a niche of abiotic conditions that can be expressed in terms of IVs for salinity, moisture and nutrient regime and that this niche can be described using a large database of vegetation plots with known association and IVs. When testing the PDF calibration accuracy, 60% of the plots were correctly classified. This suggests that the associations either do not have a strictly defined ecological niche, or that other factors than the three explanatory IVs describe the niche as well. Other potential environmental factors that may further improve distinction between associations considered here include pH, colonisation, grazing and elevation. For most of these factors, however, high spatial resolution measurements of environmental conditions were not readily available IVs for acidity (mR, ranging from 1 = acidic to 3 = alkaline) were predicted by GPR, but not used to predict association occurrence because mR related very poorly to the

spectral signal (R^2 = 0.38). However, in situations with a pronounced acidity gradient, predictions of mR may improve the discrimination of associations. Several sources offer IVs for yet other environmental factors that may be relevant for association modelling, such as light availability (Ellenberg et al., 1991) or humus content (Feldmeyer-Christe et al., 2007; Landolt, 1977). These may be within reach of remote sensing based predictions (Ecker et al., 2010).

Results also suggest that the selected associations overlapped in IV space and therefore that the occupied environmental conditions of an association were not unique. This made some associations redundant. This is particularly true for the salt tolerant associations (Table 5.3). Salt tolerant associations found in the study site are characterised by few diagnostic species, meaning that the presence or absence of a single species can lead to another classification, even though this hardly influences the average IVs of the vegetation, if at all. Partly, this may be solved by aggregating comparable associations (Verrelst et al., 2009), but we chose to maintain the original associations in order to conserve the ecological meaning and familiarity of these associations for policy makers and nature managers. Other salt tolerant associations, such as 27AA01, potentially occur at a very wide range of ecological conditions and therefore easily mix with other associations, also hampering distinction based on environmental conditions. This implies that the study site may not have been the optimal location to test this method. Indeed, earlier applications of this method, where a different selection of associations was predicted, yielded better results, with up to 85% of the plots in the national database classified correctly (Witte et al., 2007).

5.5.4. Critical evaluation of results – vegetation maps and occurrence probability maps

Both a crisp map of most likely occurring associations and an occurrence probability map per association were produced. Assessment by a local vegetation expert revealed a mismatch between predicted and observed occurrences for various associations (Jager, 2012). Some obvious errors are apparent. Since mS was positively correlated to mF, wet areas with fresh water were often assigned to high mS values. This error propagates through the vegetation maps, with a prediction of salt tolerant associations (e.g. 26AC03) around the cattle drinking ponds at the high dune complex (Fig. 5.5). Note that the association modelling currently does not take the GPR prediction standard deviation into account. Therefore, areas with high prediction standard deviation are not reflected by the association confidence map.

From the individual occurrence probability maps (supplementary data in section 5.8.3) and the distribution of most probable associations (Fig. 5.5), it is evident that some associations are predicted to be far more likely to occur than others (Fig. 5.6). While some associations are known to be more abundant in the study site (e.g. 26AA01 is very common while 06AC04 is considered rare; (Westhoff & van Oosten, 1991; Westhoff & Westra, 1981)), the distribution of predicted association occurrence defies the expected distribution based on field observations. This is caused by some associations being located in the extreme ranges

of the IV space, while GPR predicted IVs tended to be under predicted at high observed values and over predicted at low observed values. Although the predicted IVs do cover the extremes of the observed values (and even beyond those), the combination of two or three high and/or low IVs at a single pixel is apparently rare. Consequently, the IV combinations at which those associations would be expected to occur are hardly present in the IV maps, resulting in few predictions of these associations. For instance, associations 14AA02 and 14CA01 have low mS and very high mF values in the national database which translates into few predicted occurrences (Fig. 5.6). On the other hand, associations of medium rich, saline and moist environments are highly favoured (e.g., 12BA04 and 27AA01). In addition, other factors, unaccounted for in this study such as pH, elevation, soil type, light availability, succession stage and land use and management may have contributed to misfits of association predictions. However, not all of these factors are quantifiable using remote sensing techniques, nor available from other sources.

5.5.5. Suggestions for future research

The discrepancy between the distributions of associations according to expert judgement and the preexisting vegetation map and the model output, as well as the unbalanced distribution of predicted association are caused by the accumulation of two sources of error: i: mismatches in GPR-predicted IVs and ii: functional redundancy among associations. Although other studies (Kalliola & Syrjänen, 1991; Yu et al., 2006) also ran into problems when aiming to predict similar a priori defined associations, we think that this study may present a viable path towards such predictions once the challenges identified above have been solved.

Our approach used hyperspectral, high spatial resolution data, which is expensive to acquire but offers high spectral resolution. The analysis of the relevant bands revealed that most spectral regions were used in the IV predictions, although not so for each IV model individually (supplementary data in section 5.8.1). On the other hand, additional information may be available in elevation data, including elevation, orientation and slope (Ecker et al., 2010). Considering that e.g. mS in our study appears to be strongly correlated to proximity to the sea, elevation might be an especially good additional estimator for mS. The Multi Spectral Instrument (MSI) instrument aboard the Sentinel-2 satellite platform (planned launch in 2014) will provide spectral information in spectral ranges that were found relevant for IV prediction (Berger et al., 2012) on a five day basis, creating potential to predict IVs repeatedly for the same area. Currently, the backbone of this method is the database of national plots from which the PDFs of association occurrence are calculated. If and how the same database can be applied to other applications depend on the possibilities to use the plots' floristic composition to assign it to one of the desired vegetation units, as well as the feasibility of assigning IVs to the plots. Interestingly, not only IVs can be used to predict associations. A recent study computed averaged plant traits for the national plots using a plant trait database and subsequently used those traits as predictors for association occurrence probability (Douma, Aerts, et al., 2012). If remote sensing proves capable of

predicting such community mean traits, the prediction of association occurrence might greatly benefit.

5.6. Conclusion

This study evaluated a new method to map the spatial extent of a priori defined vegetation units by combining two steps which had been investigated separately in earlier studies. IVs for salinity, moisture and nutrients were predicted from VNIR and SWIR airborne remote sensing reflectance data using the novel GPR regression technique. Validation showed that extreme high and low values were systematically under- and over predicted. Following this, the ecological envelope of vegetation units was modelled as probability density functions using a large database of vegetation plots with known vegetation units and IVs. Functional redundancy in combination with GPR biases resulted in vegetation units with envelopes in the extreme ends of IVs being predicted rarely, while occurrences of vegetation units situated at frequently predicted IV combinations were over predicted. This was confirmed by expert judgement. Thus, the accumulated errors of the two modelling steps currently prevent accurate prediction of vegetation unit occurrence.

Nonetheless, the potential advantages of the method remain clear. Any combination of vegetation units that are based on floristic composition can be modelled and the method can yield both a crisp vegetation map as well as individual association occurrences. The main advantage is that vegetation unit occurrence probability is calibrated independent of the remote sensing data source and is related to vegetation properties (sensu IVs or plant traits) that are within the reach of quantitative prediction from remote sensing data. Considering the potential methods to improve IV prediction and advances to predict plant traits from remote sensing, we believe this method provides a promising roadmap to advance remote sensing sciences further into the domain of a priori defined vegetation units.

5.7. Acknowledgements

We thank Pieter Slim (Alterra), Stephan Hennekens (Alterra) and Rijkswaterstaat for providing vegetation plot data and Anna Schmidt (Alterra) for providing the AHS remote sensing image. T.H. Jager kindly volunteered to review the results. Han Runhaar (KWR), Martin de Haan (KWR) and Bob Douma (WUR) are acknowledged for their classification of the vegetation data and critical evaluation of the results. Hella van Asperen is acknowledged for fieldwork assistance. This research was conducted as part of the joint research program of KWR Watercycle Research Institute. Four anonymous reviewers provided valuable suggestions that greatly improved this work.

5.8. Supplementary data

5.8.1. Band relevance in GPR modelling

An important feature of Gaussian Process Modelling (GPR) is that it provides insight in the relevance of each predictor variable when fitting a model. However, these values tend to be unstable when a different calibration set is used. To reliably estimate the relevance of the AHS bands in the indicator value predictions, we fitted 30 GPR models for each indicator value, using a random 80% of the local plots as calibration. Fig. 5.7 indicates how often each band was the most relevant, second, third or fourth most relevant band.

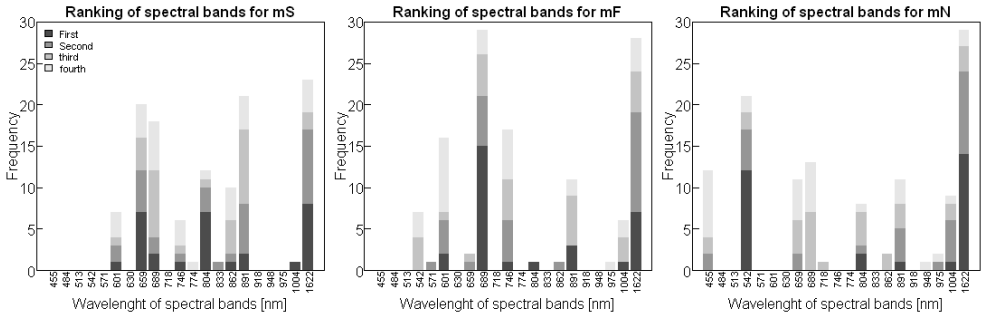

Figure 5.7 Relevance of each spectral band, for fitting the GPR models. The frequency is shown of each band being the most relevant, second, third of fourth most relevant band of the model.

5.8.2. Prediction standard deviation GPR models

In addition to a mean prediction, Gaussian process regression also provides an associated confidence measure for each pixel, which is shown in Fig. 5.8. The values shown here represent the standard deviation in the prediction, meaning that lower values signify a more confident prediction.

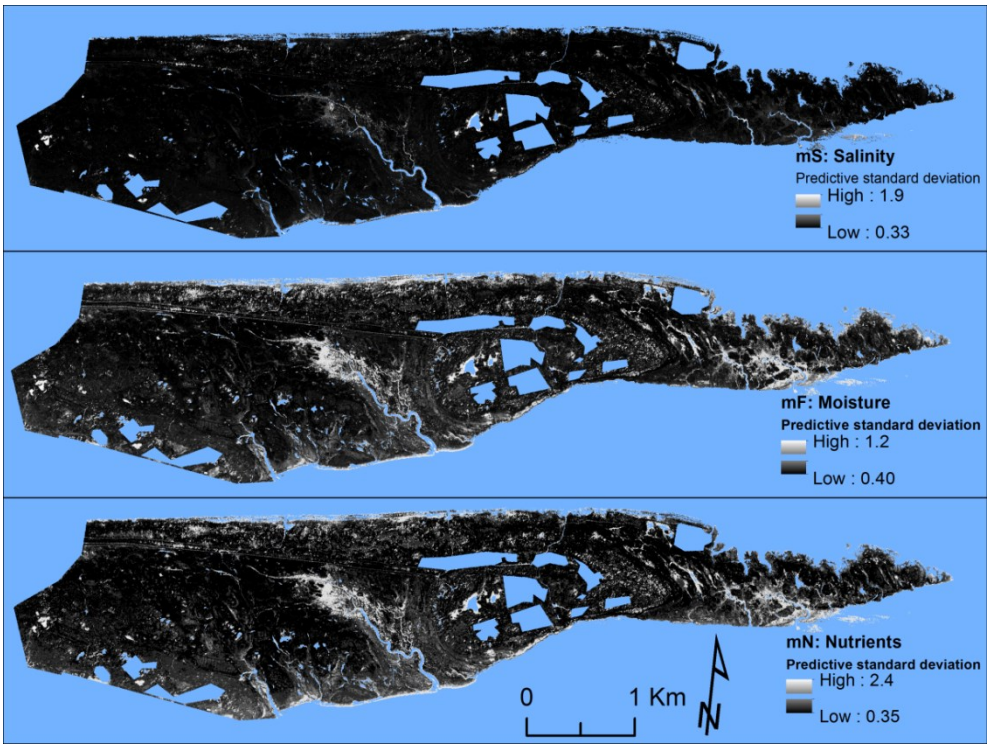

Figure 5.8 Per pixel confidence level of the GPR IV predictions.

5.8.3. Occurrence probabilities of the associations and comparison with the pre-existing vegetation map

The occurrence probability was calculated for each association by applying the probability density functions (PDFs) to the GPR-estimated indicator values. The resulting probabilities are visualised in figures 5.9 – 5.12 in the left column.

The right column of figures 5.9 – 5.12 shows the distribution of the associations according to the pre-existing vegetation map. The area-share of the association in question within each polygon area is indicated on an ordinal scale. The mapping extent of the pre-existing vegetation map is hatched. Note that this extent is smaller than the study site. Hence, the comparison only holds true for the overlapping (hatched) area. Not all 19 associations are included in the pre-existing vegetation map, because the association in question was not found in the mapping extent of this map. If this is the case, the map in the right column is blank.

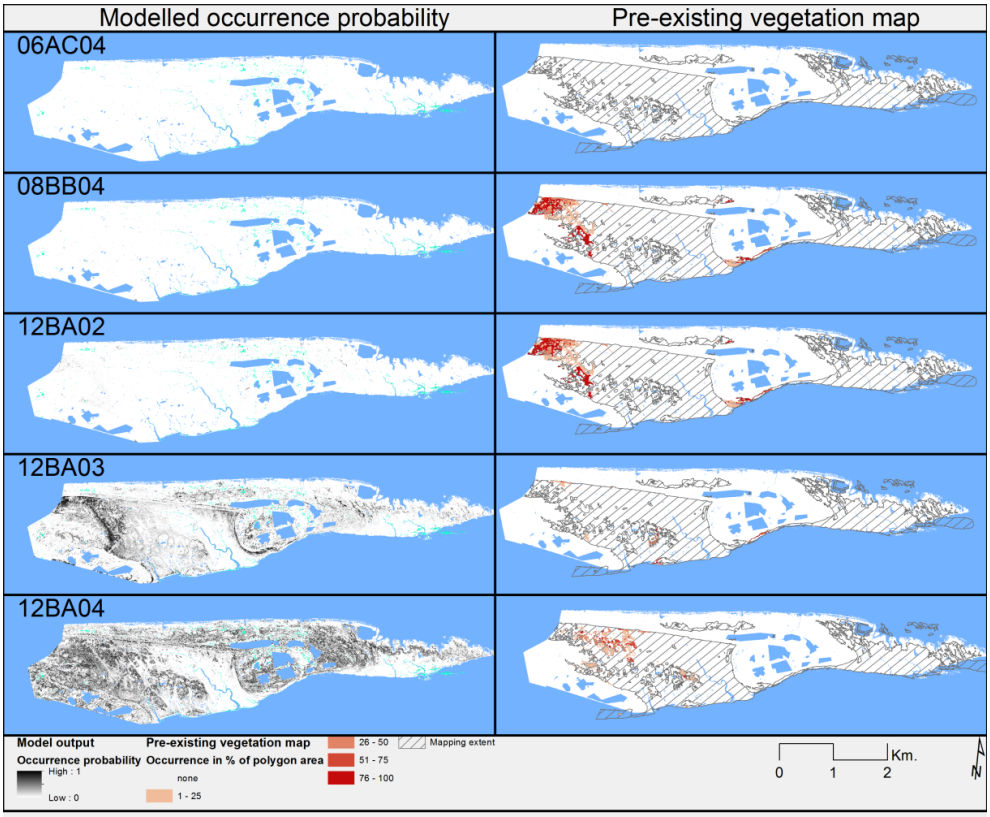

Figure 5.9 Modelled occurrence probability and occurrence following the pre-existing vegetation map for associations 06AC04 – 12BA04.

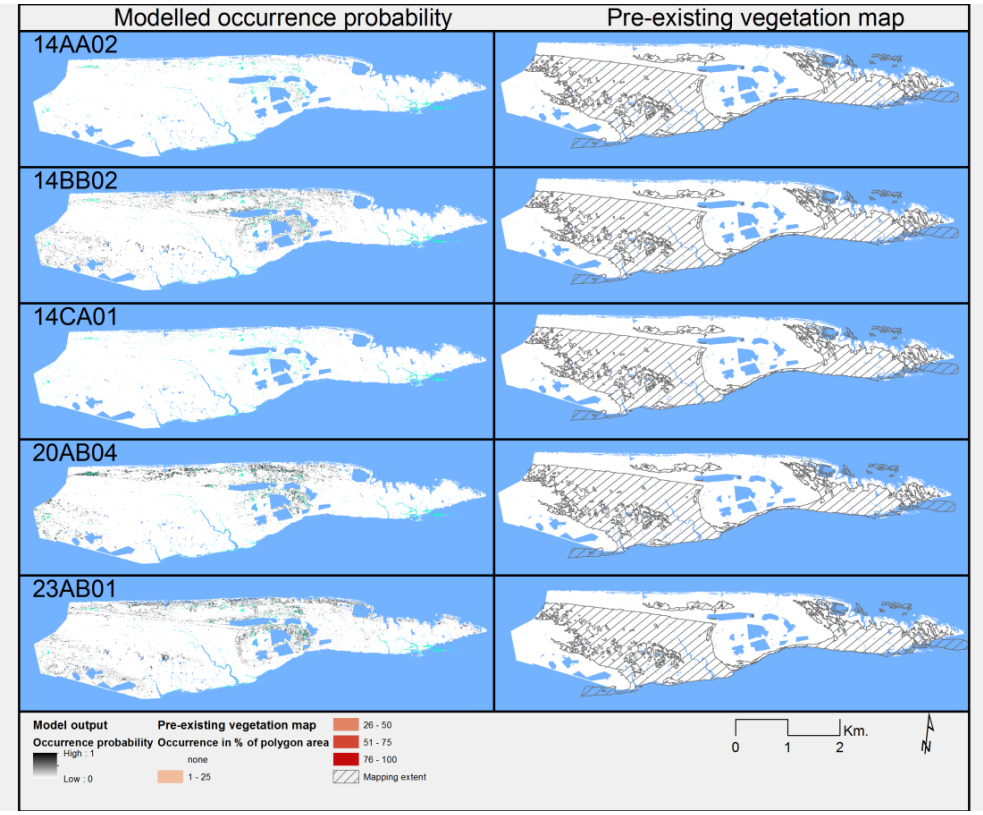

Figure 5.10 Modelled occurrence probability and occurrence following the pre-existing vegetation map for associations 14AA02 – 23AB01.

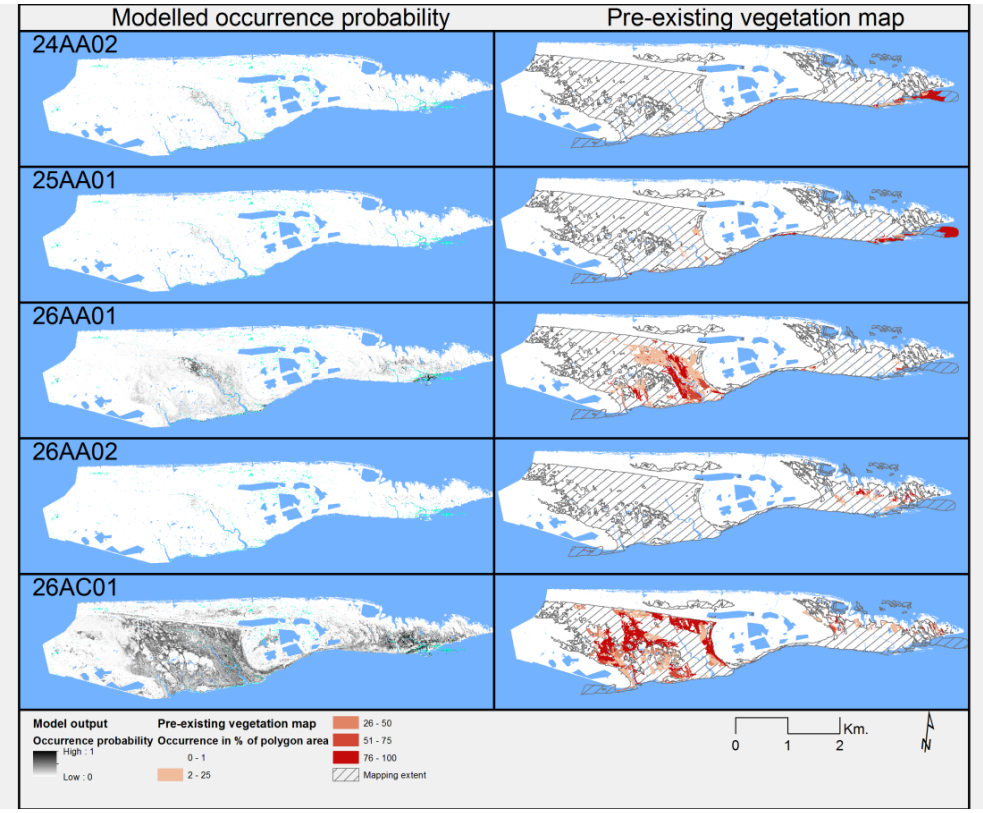

Figure 5.11 Modelled occurrence probability and occurrence following the pre-existing vegetation map for associations 24AA02 – 26AC01.

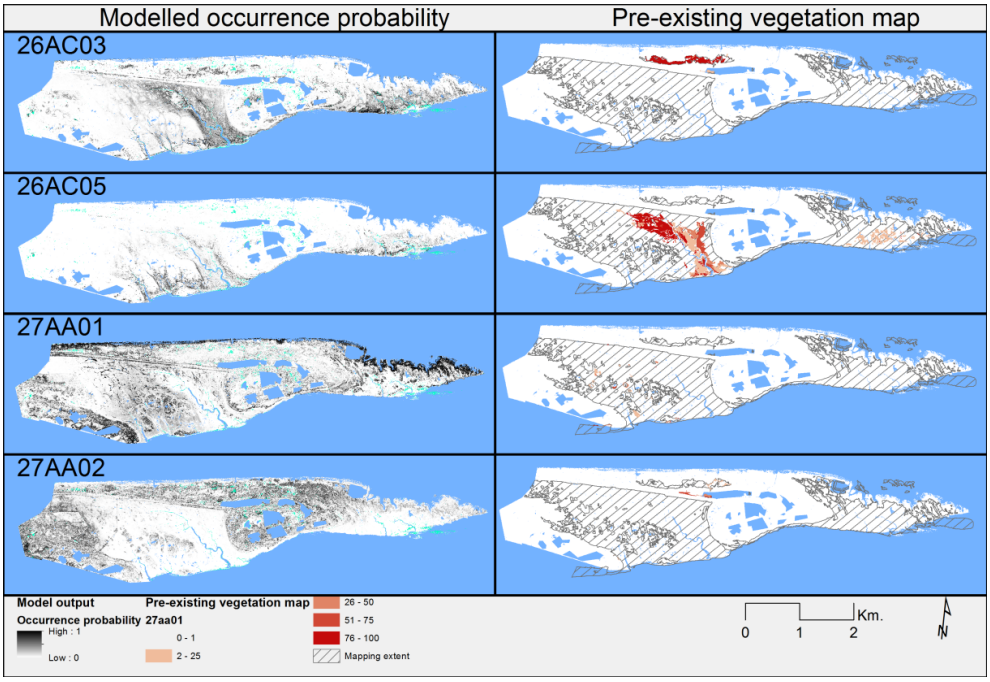

Figure 5.12 Modelled occurrence probability and occurrence following the pre-existing vegetation map for associations 26AC03 – 27AA02.